CLAIMS

What is Claimed is:

1. A method of dynamically balancing load in a system of servers, comprising:

a) monitoring for servers that are able to respond to requests directed at the system;

b) determining a performance metric for servers discovered by said monitoring for the servers;

c) maintaining a table comprising said performance metric for said discovered servers; and

d) in response to receiving a request, routing said request to a selected server in the system of servers based on said performance metric, wherein the system of servers comprises the discovered servers.

2. The method of Claim 1, further comprising:

determining a load on ones of the servers in the system of servers.

3. The method of Claim 2, further comprising:

determining a stress factor for a given server based on the performance metric of the given server and the load on the given server.

4. The method of Claim 1, further comprising:

determining a stress factor for ones of the servers in the system of servers based on the performance metrics.

5. The method of Claim 1, wherein the performance metric is a response time.

6. The method of Claim 1, wherein the performance metric is a response time when the servers discovered by said monitoring are unloaded.

7. The method of Claim 1, further comprising:

periodically reevaluating said performance metric for the servers in the system of servers.

8. A method of dynamically balancing load, comprising:

a) dynamically discovering servers that are able to respond to requests directed at a system;

b) determining a response time of each of the discovered servers;

c) calculating stress factors for each of the discovered servers, based in part on said response time;

d) receiving a request to the system;

e) determining a server in the system to route the request to based on the stress factors, wherein the system comprises the discovered servers; and

f) routing said request to said server in the system determined in said e).

9. The method of Claim 8, wherein said b) comprises determining a response time for each of the discovered servers to a request.

10. The method of Claim 8, wherein said b) comprises determining a response time for each of the discovered servers to a database query.

11. The method of Claim 8, wherein said c) comprises calculating the stress factor for each of the discovered servers, based on said response time and a load for each of the discovered servers.

12. The method of Claim 8, wherein:

said b) further comprises determining a response time of servers not discovered in said a);

said c) comprises calculating stress factors for each of the servers not discovered in said a), wherein the system further comprises the servers not discovered in said a).

13. The method of Claim 8, wherein said servers not discovered in said a) are reported to a load-balancing agent in a configuration file.

14. A system for balancing load, comprising:

a plurality of back-end servers that are able to service requests to the system;

a front-end server having a load balancing agent comprising a table, wherein said front-end server receives requests that are forwarded to said back-end servers, and wherein said load balancing agent is operable to:

monitor for back-end servers that are able to service requests to the system;

determine a performance metric for the back-end servers discovered by the monitoring; and

determine a server of said back-end servers to route a request to based on the performance metric.

15. The system of Claim 14, wherein said load balancing agent is further operable to determine a load on a given back-end server.

16. The system of Claim 14, wherein said load balancing agent is further operable to determine a stress factor for ones of the back-end servers.

17. The system of Claim 16, wherein the stress factor for a given one of the back-end servers is based on the performance metric and the load on a given of the given one of the back-end servers.

18. The system of Claim 17, wherein said load balancing agent is able to determine which server of said back-end servers to route a request to based on the stress factor.

19. The system of Claim 14, wherein the performance metric is a response time.

20. The system of Claim 17, wherein said load balancing agent is able to include back-end servers that the load balancing agent did not discover in the determination of which server to route the request to.